

Análise Fatorial Exploratória

TUANY DE PAULA CASTRO

Análise Fatorial Exploratória

O que é?

Técnica para reduzir o número de variáveis de uma base de dados

Quando usar?

Para resumir informações e mensurar fenômenos latentes

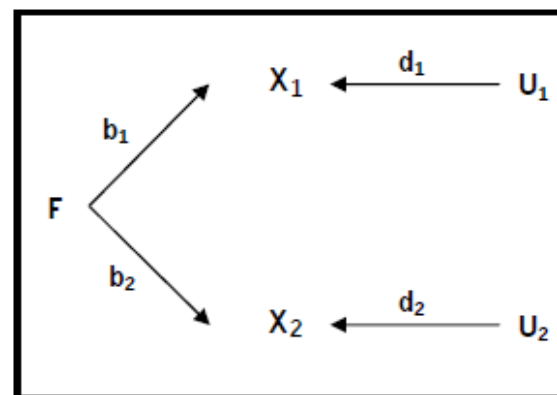
Finalidades e Aplicações

O problema consiste em obter combinações lineares das variáveis:

- Estas combinações lineares são chamadas fatores;
- Os fatores são variáveis latentes que dão origem às variáveis reais;
- Os fatores também variam entre os indivíduos, porém não podem ser medidos ou observados;
- A existência dessas variáveis hipotéticas é uma questão em aberto.

Figura 1

Modelo das vias para duas variáveis, modelo de um fator comum



Fonte: Asher, 1983

Finalidades e Aplicações

Diferença entre Análise Fatorial e Análise de Componentes Principais:

“ Se você estiver interessado numa solução teórica não contaminada por variabilidade de erro, a análise fatorial deve ser sua escolha. Se você quiser simplesmente um resumo empírico do conjunto de dados, a análise de componentes principais é uma escolha melhor.” (Tabachinick e Fidell, 2007)

“A análise de componentes principais é em geral preferida para fins de redução de dados, enquanto a análise fatorial é em geral preferida quando o objetivo da pesquisa é detectar a estrutura dos dados da modelagem casual” (Garson, 2009)

Raciocínio estrutural

Procedimento	O que deve ser observado
Verificar a adequabilidade da base de dados.	Nível de mensuração das variáveis, tamanho da amostra, razão entre o número de casos e a quantidade de variáveis e o padrão de correlação entre as variáveis.
Determinar a técnica de extração e o número de fatores a serem extraídos.	O tipo de extração (componentes principais, fatores principais, máxima verossimilhança, mínimos quadrados, mínimos quadrados ponderados)
Decidir o tipo de rotação dos fatores.	Se for ortogonal (varimax, quartimax, equamax), se for oblíqua

Critérios para o uso

Nível de mensuração	Variáveis discretas e contínuas
Amostra	Amostras mínimas de 50 casos; razão mínima de 5 casos para uma variável.
Correlação	Maior parte dos coeficientes de correlação deve apresentar valores acima de 0,30.
KMO	Quanto maior, melhor, tendo 0,50 como o patamar mínimo de adequabilidade (Hair t al, 2006).
BTS	$p < 0,05$

Extração dos fatores

Tipo de extração	Determinar a técnica de extração dos fatores (<i>principal components, principal factors, image factoring; maximum likelihood factoring; alpha factoring; unweight least square; generalized least square</i>)
Regra de Kaiser	Devem ser extraídos apenas os fatores com valor do <i>eigenvalue</i> acima de 1
Scree test	Analisar graficamente a dispersão do número de fatores até que curva da variância individual de cada fator se tornar horizontal ou sofrer uma queda abrupta
Variância acumulada	> 60%
Razão teórica	O pesquisador deve justificar teoricamente como as variáveis se relacionam com os fatores extraídos.

Extração dos fatores: métodos mais usados

Principais eixos fatoriais

- Busca o menor número de fatores que podem explicar a variância comum (correlação) de um conjunto de variáveis;
- Mais capaz de reconhecer fatores fracos;
- Mais recomendado em caso de dados com distribuição não-normal.

Máxima verossimilhança

- O modelo estima as cargas e as comunalidades que maximizam a probabilidade da matriz de correlação observada;
- Assintoticamente eficiente (permite avaliação da qualidade do ajuste, testes de significância dos fatores e obtenção de intervalos de confiança);
- Mais recomendado em caso de dados com distribuição normal.

Exemplo 1

Considere um conjunto com 9 variáveis cognitivas de 4175 alunos de um exame de entrada num colégio. As variáveis são:

- Definitions
- Arithmetical Problems
- Classification
- Artificial Language
- Antonyms
- Number Series Completion
- Analogies
- Logical Inference
- Paragraph Reading

Exemplo 1

Verificando **critérios para uso**:

- Tamanho amostral: $N = 4175$
- Razão entre número de observações e quantidade de variáveis:
 $4175/9 = 463,89$
- KMO = 0,90 (ótimo)

Os dados estão adequados para análise fatorial.

Exemplo 1

Extração dos fatores sem rotação:

	Eixos Principais		Máxima Verossimilhança	
	P1	P2	M1	M2
Definitions	0,77	-0,01	0,60	0,48
Arithmetical_Problems	0,64	0,40	0,45	0,50
Classification	0,73	-0,10	0,57	0,44
Artificial_Language	0,65	0,12	0,42	0,55
Antonyms	0,81	-0,27	0,77	0,27
Number_Series_Completion	0,58	0,52	0,35	0,57
Analogies	0,74	-0,06	0,55	0,50
Logical_Inference	0,75	0,01	0,64	0,39
Paragraph_Reading	0,81	-0,36	1,00	-0,02
Autovalor	4,72	0,66	4,60	0,50
Variância Explicada	0,52	0,08	0,51	0,06
Variância acumulada	0,52	0,60	0,51	0,57

Exemplo 1

Comunalidades:

	Eixos Principais	Máxima Verossimilhança
Definitions	0,61	0,56
Arithmetical_Problems	0,60	0,45
Classification	0,57	0,52
Artificial_Language	0,48	0,43
Antonyms	0,73	0,68
Number_Series_Completion	0,61	0,41
Analogies	0,62	0,53
Logical_Inference	0,56	0,54
Paragraph_Reading	0,94	0,65

Exemplo 1

Conclusões:

- **Eixos Principais:**

- Com dois fatores, explica-se 60% da variabilidade dos dados;
- Comunalidades maiores;
- Primeiro fator geral;
- Segundo fator relacionado à habilidade *Number Series Completion*.

- **Máxima Verossimilhança**

- Com dois fatores, explica-se 57% da variabilidade dos dados;
- Primeiro fator relacionado às habilidades de linguagem;
- Segundo fator relacionado às habilidades com números.

Métodos de rotação de fatores

- ❑ É um método matemático que rotaciona os eixos no espaço geométrico;
- ❑ Esse método torna mais fácil determinar quais variáveis são carregadas em quais componentes;
- ❑ Objetivo principal: tornar o resultado empírico encontrado mais facilmente interpretável, conservando as suas propriedades estatísticas.

Métodos de rotação de fatores

“Execução aritmética para obter um novo conjunto de cargas fatoriais de um determinado conjunto.” McDonald (1985, p. 40)

“Um procedimento no qual os fatores são girados na tentativa de conseguir uma estrutura mais simples.” Bryant and Yarnold (1995, p. 132)

“Qualquer um dos vários métodos em análise fatorial pelo qual o pesquisador tenta relacionar fatores calculados com entidades teóricas. Isso é feito de formas diferentes, dependendo se os fatores são considerados correlacionados (oblíquos) ou não correlacionados (ortogonais).” Vogt (1993, p. 91)

“Em análise fatorial ou de componentes-principais, é feita a rotação dos eixos fatoriais (dimensões) identificados inicialmente para obter fatores mais simples e interpretáveis.” Yaremko, Harari, Harrison and Lynn (1986).

Métodos de rotação de fatores

Existem dois tipos de rotação:

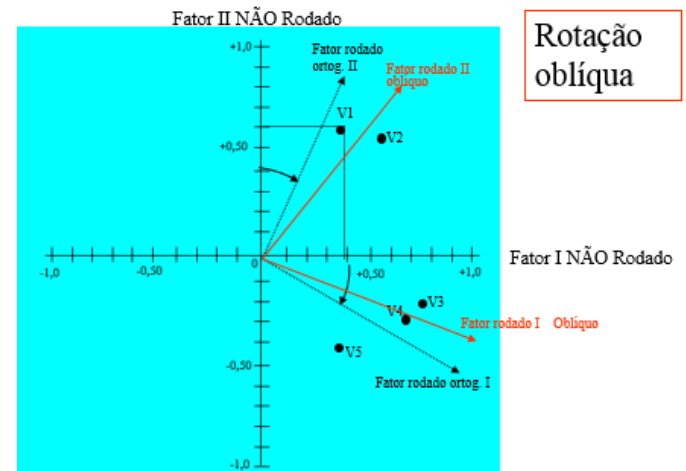
- ❑ Ortogonal: supõe fatores independentes. Mais fácil interpretação:
 - Varimax
 - Quartimax
 - Equamax

- ❑ Oblíqua: fatores podem ser correlacionados, porém mais difícil interpretação:
 - Oblimin
 - Promax

Métodos de rotação de fatores

Considerações:

- A rotação não altera a posição das variáveis no espaço dos fatores, ou seja, as correlações entre variáveis são preservadas;
- Uma rotação **ortogonal** altera as variâncias dos fatores, porém os mantém não correlacionados e preserva as communalidades;
- Uma rotação **oblíqua** permite que os fatores sejam correlacionados se isso produzir uma estrutura mais simples para análise. No entanto, a interpretação se torna mais difícil, pois se deve interpretar os fatores paralelamente e não um a um.



Métodos de rotação de fatores

Como escolher entre rotação ortogonal e oblíqua?

A literatura propõe que se faça uma rotação oblíqua e então analise as correlações dos fatores.

Se a rotação oblíqua mostrar que os fatores são fracamente correlacionados (correlações menores do que 0,30), pode-se utilizar a rotação ortogonal. Se os fatores, por outro lado, estiverem muito correlacionados, pode-se extrair menos fatores numa rotação ortogonal ou, alternativamente, usar os resultados oblíquos.

Métodos de rotação de fatores

Varimax

- ✓ Opção de rotação mais comum;
- ✓ Procura minimizar o número de variáveis que apresentam altas cargas em cada fator.

Quartimax

- ✓ Procura minimizar o número de fatores necessários para explicar cada variável;
- ✓ Geralmente gera um fator geral sobre o qual a maioria das variáveis são carregadas em um grau alto ou médio.

Equamax

- ✓ Compromisso entre os critérios Varimax e Quartimax.

Métodos de rotação de fatores

Oblimin

- ✓ Método oblíquo padrão (fatores podem ser correlacionados);
- ✓ Autovalores mais altos, mas interpretabilidade diminuída.

Promax

- ✓ Método oblíquo alternativo, computacionalmente mais rápido.

Métodos de rotação de fatores

Como escolher dentre os tipos particulares de rotação?

Deve-se escolher o tipo de rotação que leve à estrutura mais simples. Uma estrutura simples é alcançada quando obtém-se as seguintes características:

- Cada variável deve produzir pelo menos uma carga zero em algum fator;
- Cada fator deve ter pelo menos tantas cargas zeros quanto o número de fatores;
- Cada par de fatores deve ter variáveis com cargas significativas em um e iguais a zero no outro;
- Cada par de fatores deve ter uma grande proporção de cargas zero em ambos fatores (se houver quatro ou mais fatores totais);
- Cada par de fatores deve ter apenas algumas variáveis complexas.

Métodos de rotação de fatores

Questões:

- ☐ **O que é uma carga zero?** Qualquer valor que fique entre -0,10 e +0,10.
- ☐ **O que é uma carga significativa?** Em uma amostra com pelo menos 100 pessoas, cargas acima de 0,30 podem ser consideradas significativas.
- ☐ **O que são variáveis complexas?** São variáveis com carga acima de 0,30 em mais de um fator.

Exemplo 2

Considere um conjunto com as seguintes variáveis sobre personalidade:

- *Social extraversion (extroversão social)*
- *Ascendance (domínio)*
- *Thinking extraversion (extroversão de pensamento)*
- *Rhathymia (despreocupação)*
- *General Activity (atividades gerais)*
- *Lack of agreeableness (falta de amabilidade)*
- *Lack of cooperativeness (falta de cooperatividade)*
- *Lack of objectivity (falta de objetividade)*
- *Nervousness (nervosismo)*
- *Inferiority feelings (sentimentos de inferioridade)*
- *Cyclic tendencies (tendências cíclicas)*
- *Depression (depressão)*

Exemplo 2

Matriz de correlação para os três fatores da análise fatorial exploratória com rotação *oblimin*.

Fator	1	2	3
1	1,00	-0,08	0,08
2	-0,08	1,00	0,00
3	0,08	0,00	1,00

Como nenhuma correlação ficou acima de 0,30, a solução para o problema pode ser ortogonal.

Exemplo 2

Resultados para três rotações ortogonais:

<i>Trait</i>	Varimax Rotation			Quartimax Rotation			Equamax Rotation		
	Factor			Factor			Factor		
	1	2	3	1	2	3	1	2	3
social extraversion	-0.135	0.665	-0.118	-0.146	0.662	-0.122	-0.129	0.666	-0.115
ascendancy	-0.109	0.548	-0.098	-0.118	0.545	-0.100	-0.104	0.549	-0.095
thinking extraversion	-0.072	-0.026	0.530	-0.072	-0.024	0.530	-0.073	-0.028	0.530
rhythmic	0.391	0.606	0.206	0.381	0.613	0.203	0.396	0.602	0.209
general activity	-0.219	0.680	-0.070	-0.230	0.676	-0.073	-0.213	0.682	-0.067
lack of agreeableness	0.119	0.540	0.197	0.110	0.543	0.194	0.124	0.538	0.199
lack of cooperativeness	0.466	0.029	0.060	0.465	0.037	0.060	0.466	0.025	0.060
lack of objectivity	0.614	0.044	-0.187	0.614	0.053	-0.188	0.615	0.039	-0.187
nervousness	0.759	-0.171	-0.021	0.762	-0.158	-0.020	0.758	-0.177	-0.021
inferiority feelings	0.677	-0.469	0.101	0.685	-0.457	0.104	0.673	-0.476	0.099
cyclic tendencies	0.785	0.106	-0.005	0.783	0.119	-0.006	0.785	0.099	-0.004
depression	0.784	-0.227	-0.076	0.788	-0.214	-0.076	0.782	-0.234	-0.077

Resultados equivalentes. Rotação escolhida: varimax.

Exemplo 2

Resultados para duas rotações oblíquas:

<i>Trait</i>	Oblimin Rotation			Promax Rotation		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
social extraversion	-	0.666	-	-	0.672	-
	0.169		0.139	0.200		0.125
ascendance	-	0.549	-	-	0.553	-
	0.137		0.115	0.163		0.103
thinking extraversion	-	-0.015	0.523	-	-0.014	0.492
	0.070			0.057		
rhathymia	0.360	0.597	0.227	0.336	0.575	0.322
general activity	-	0.685	-	-	0.695	-
	0.254		0.098	0.284		0.100
lack of agreeableness	0.092	0.540	0.198	0.072	0.531	0.241
lack of cooperativeness	0.464	0.016	0.097	0.462	-0.008	0.180
lack of objectivity	0.611	0.022	-	0.603	-0.008	-
			0.138			0.019
nervousness	0.767	-0.194	0.043	0.771	-0.231	0.172
inferiority feelings	0.700	-0.488	0.163	0.721	-0.521	0.258
cyclic tendencies	0.778	0.082	0.057	0.771	0.042	0.204
depression	0.794	-0.252	-	0.800	-0.290	0.123
			0.010			

Exemplo 2

Conclusão:

- ❖ Conforme avaliado anteriormente, não houve diferença significativa entre as interpretações obtidas com rotação ortogonal e oblíqua, optando-se pela rotação ortogonal devido à interpretação mais simples.
- ❖ Dentre as rotações ortogonais, as rotações foram equivalentes, optando-se pela varimax.

Exercícios

1) Foi aplicado um questionário para avaliar as seguintes características de um grupo de pacientes de um hospital: alegria, boa disposição, animação, preocupação, receio e ansiedade. Uma análise fatorial por Eixos Principais e rotação Varimax foi aplicada a esses dados e foram determinados os seguintes fatores indicados abaixo:

$$\textit{Alegria} = 0,12 * F_1 + 0,85 * F_2$$

$$\textit{Boa disposição} = 0,07 * F_1 + 0,74 * F_2$$

$$\textit{Animação} = 0,18 * F_1 + 0,67 * F_2$$

$$\textit{Preocupação} = 0,93 * F_1 + 0,21 * F_2$$

$$\textit{Receio} = 0,77 * F_1 + 0,05 * F_2$$

$$\textit{Ansiedade} = 0,62 * F_1 + 0,08 * F_2$$

Exercícios

Com base nesses resultados:

- (a) Calcule os autovalores dos fatores. Comente.
- (b) Qual a variância explicada pelos fatores?
- (c) Obtenha as comunalidades das variáveis.
- (d) Faça um gráfico que represente os fatores e as variáveis. Interprete os fatores.

Exercícios

2) O conjunto *dados_animais.csv* contém os resultados de teste aplicado em 154 pessoas acerca da opinião sobre experimentações e direitos de animais. Cada participante da pesquisa avaliou 28 afirmações em cinco categorias: 1 – discordo totalmente, 2 – discordo, 3 – não tenho opinião formada, 4 – concordo e 5 – concordo totalmente. No arquivo *pesquisa_animal.docx*, encontram-se as afirmações do questionário.

- (a) Avalie a adequabilidade do uso de uma análise fatorial para os dados.
- (b) Identifique, por meio do Scree Plot, o número ideal de fatores para o modelo (extração por componentes principais).
- (c) Utilizando rotação Varimax, compare os resultados obtidos com as extrações por componentes principais, eixos principais e máxima verossimilhança. Interprete os fatores encontrados.

Exercícios

3) Em *dados_mc.csv*, encontram-se dados referentes a uma avaliação nutricional de produtos vendidos pela rede MC Donald's. Queremos identificar as variáveis latentes que caracterizam os produtos nutricionalmente. Para tanto:

- (a) Verifique se um modelo de análise fatorial é adequado.
- (b) Por meio dos critérios de Kaiser, variância acumulada e Scree Plot, escolha um número de fatores para o modelo.
- (c) Utilizando a extração por Componentes Principais, verifique qual o melhor método de rotação para a análise (varimax ou oblimin). Justifique.
- (d) Com o método de rotação definido, interprete os fatores.

Exercícios

4) Os dados *Wine_data.csv* correspondem a 12 características avaliadas em 1599 vinhos. Tem-se interesse em estudar fatores latentes a essas características.

- (a) Verifique se a análise fatorial é adequada para o estudo.
- (b) Defina um número de fatores para o modelo.
- (c) Utilizando o método de Componentes Principais, avalie qual tipo de rotação é mais indicado aos dados (*oblimin* ou *varimax*).
- (d) Indique a variável mais e a menos explicada pelos fatores em estudo.
- (e) Interprete os fatores obtidos.
- (f) Compare os resultados obtidos com as extrações por eixos principais e máxima verossimilhança.

Referências

- BISQUERRA, R; CASTELLA, J.; VILLEGAS, F. Introdução à estatística: enfoque informático com o pacote estatístico SPSS. Porto Alegre: Artmed, 2007.
- DANCEY, Cristine P; REIDY, John. Estatística sem Matemática Para Psicologia. 3 edição. Porto Alegre: Artmed, 2006.
- HAIR, J.; ANDERSON, R.; BLACK, W. Análise multivariada de dados. 6 ed. Reimp. Porto Alegre: Bookman, 2009.
- JOHNSON, R. and WICHERN, D. Applied Multivariate Statistical Analysis. Sixth edition, Wisconsin, Pearson, 2007.